CLIMATE SIGNALS FOR ENHANCED RUNOFF FORECASTING IN WESTERN U.S. REGIONS

Levi Brekke, Hydraulic Engineer, U.S. Bureau of Reclamation, Denver, CO, lbrekke@do.usbr.gov
Jon Medina, Meteorologist, U.S. Bureau of Reclamation, Denver, CO, jmedina@do.usbr.gov
David Raff, Hydraulic Engineer, U.S. Bureau of Reclamation, Denver, CO, draff@do.usbr.gov
Shaleen Jain, Staff Scientist, NOAA Climate Diagnostic Center, Boulder, CO, Shaleen.Jain@noaa.gov

Abstract: Teleconnection climate signals have been selected and evaluated for improving April-July runoff forecasting in Western U.S. basins. Signal selections occur at lead-times varying from coincident winter to prior summer for enhanced short-lead forecasts and experimental longer-lead forecasts. Selections are seasonal conditions observed in the North Pacific 700mb and (500-700)mb geopotential height fields. Shorter-lead signals are generally located over the Northeast Pacific and have 0.5 to 0.8 correlations with runoff. Longer-lead signal locations are in the Subtropical and Northwest Pacific, and have 0.4 to 0.6 with runoff. Selection of signals relies on analysis of signal relations to subsequent April-July runoff, coincident Pacific Ocean surface temperatures and winter atmospheric circulation over the Western U.S. This leads to a candidate signal set that is evaluated via comparative forecasting with and without the teleconnection signal information, using an operational technique for shorter-lead forecasting, and an experimental technique for longer-lead. Results indicate that signals offer additive value relative to current information sets used in January and February forecasting (i.e. antecedent precipitation, antecedent runoff and snowpack). Skillful longer-lead forecasting in October, November, and December is also possible for many of the basins analyzed. A tool was developed to assist deployment of these results and procedures to operational forecasting centers (U.S. Bureau of Reclamation's Pacific Northwest Region - River and Reservoir Operations Group (USBR PN6200), CA Department of Water Resources – Flood Management Division (CA DWR), U.S. Department of Agriculture's Natural Resources Conservation Service - National Water and Climate Center (NRCS), and National Weather Service CA-NV River Forecast Center (CNRFC)), a tool has been developed to assist signal data retrieval and management, and to enable longer-lead forecasting.

INTRODUCTION

For water resources management in the Western United States, hydroclimatic forecasts provide key information for seasonal to annual planning processes, including reservoir operations scheduling, hydropower wholesaling, and water transfers planning. Calendars of these planning processes differ by when plans are issued and for what horizon. The anticipated water supply arriving, largely as snowmelt during April-July, is a common centerpiece to these planning processes.

Statistical forecasts of April-July runoff are made available to planning agencies starting in January of each calendar year. Forecasts are developed by various operational centers (e.g., NRCS, CNRFC, CA DWR, USBR PN-6200), and are based on season-to-date hydroclimate information (e.g., precipitation since October, snowpack accumulated near the time of forecast issue, antecedent Autumn runoff as a soil-moisture proxy).

A potential source of additional forecast information may be teleconnection signals. Teleconnections are defined here as the coupled variations between geographically and/or temporally separated climate phenomena. Time-separate teleconnections are relevant for hydroclimate forecasting and the subject here. The potential role of teleconnections to improve season runoff forecasting in the Western United States has been well documented. Much of focus has been on teleconnections involving lead signals describing El Nino/Southern Oscillation (ENSO) and Pacific Decadal Oscillation (PDO) conditions (**Redmond and Koch 1991, Souza and Lall 2003**). ENSO/PDO-based runoff forecasts have been shown to economically benefit Pacific Northwest hydropower and reservoir operations (**Hamlet et al. 2002**).

A limitation of using ENSO/PDO-based runoff forecasts is that many basins in Western U.S. regions (e.g., Northern California, Intermountain West) do not exhibit consistent relation to antecedent ENSO/PDO conditions. One alternative approach is to use climate signals embedded in large-scale ocean atmospheric features that have relationships to basin-specific runoff variations (**Grantz et al 2005**). This idea is supported by past findings of how small variations in large-scale atmospheric patterns can lead to large changes in surface climate (**Yarnal and Diaz 1986, Clark and Serreze 1999**). This manuscript presents a project designed to test the applicability of using embedded large-scale atmospheric features as additive information for April-July forecasts of water supply in west-coastal United States regions (WA, OR, CA, NV, ID, and western WY and MT). The climate signals are upwind atmospheric anomalies over the North Pacific Ocean. This manuscript presents, specifically, the methodologies used for signal identification and information assessment. Seventeen basins were analyzed in total; summary results on two basins are presented here (i.e. Yakima Basin in WA, Truckee Basin in CA/NV). Description and deployment of an experimental forecasting tool based on these signals selections is also presented.

METHODOLOGY PART I: SIGNAL IDENTIFICATION

Teleconnection signals were considered from an upwind atmospheric area relative to the Western United States (i.e. within 20-60N and 100-255E), which represents a general region where West Coast U.S. storms originate and propagate. It is presumed that atmospheric pressure structures (i.e. geopotential height conditions at 700mb and 500mb, represented historically by monthly NCEP/NCAR Reanalysis data provided by the NOAA-CIRES Climate Diagnostics Center (CDC), Boulder, Colorado, USA, www.cdc.noaa.gov/, sampled at 5 degree spatial resolution) will exhibit anomalous conditions as these climatic influences propagate through the region to affect Western U.S. hydrocliomate In addition to identifying climate signals in the distributed North Pacific atmosphere, signals were also considered from conventional monthly ENSO-indices (e.g. NINO*, SOI, MEI).

The first step of identifying teleconnection signals utilized correlation analysis between a basin's April-July runoff and an antecedent signal *scheme* defined by (a) variable, (b) season, and (c) aggregation. The variable was either an ENSO-index, a location-specific atmospheric geopotential height at the 700 and 500 mb pressure level (Z_{700} and Z_{500}, respectively), or a location-specific difference between these two heights ($\Delta_{(500\text{-}700)}$). The season was allowed to be two to six months, occurring with up to 1 year lead-time. Aggregation was either season-mean or trend, defined as the value of the last month minus the value of the first month. The correlation coefficient,*r*, between each *scheme*-dependent signal and April-July runoff was computed from 1950-1998 for Pacific Northwest basins and 1952-2000 for Mid-Pacific basins. Runoff data were provided for CA/NV/Southern-OR basins by CNRFC, and for WA/ID/MT/WY basins by USBR PN-6200. If *r* passed a 99% significance test, the signal was kept for further consideration. For *schemes* involving location-dependent variables (i.e., Z_{700}, Z_{500}, and $\Delta_{(500\text{-}700)}$), the *scheme* might produce significant correlations involving numerous locations. In these cases, an optimal correlation location was identified using spatial mapping of the variable-runoff correlation as it varied by Reanalysis-location.

Once a subset of candidate teleconnections signals were identified, based on *r*, a further selection criteria was imposed for North Pacific atmosphere signals based on whether the chosen signal could be physically related to the runoff. In most cases, this question has yet to be completely answered. However, a first step in this evaluation was performed by checking whether the signal's seasonally persistent atmospheric condition was plausible. The expectation, given plausibility, is that the signal would have a coincidental relationship with sea surface temperature variations in the Pacific Basin. The assumption is that seasonal persistence in the atmosphere must be linked to seasonal persistence at the ocean surface, which generally forces the atmosphere on 2- to 6-month time scales. This expectation was tested using an additional correlation analysis, where the atmospheric signal is the "basepoint" and the distributed variable is sea surface temperature anomaly during the signal season (i.e. Kaplan SST dataset provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, www.cdc.noaa.gov/, sampled at 5 degree spatial resolution).

METHODOLOGY PART II: INFORMATION ASSESSMENT

After identifying basin-specific sets of ENSO- and Mid-Latitude-Atmosphere teleconnection signals, the potential information value of each signal set was assessed. Forecast simulations for April-July runoff were conducted at longer-lead issue times (i.e. October 1st, November 1st, December 1st) and at current, operational issue times (i.e. January 1st, February 1st, … April 1st). Forecasts were generated using two candidate information sets: (a) default information (which includes season-to-date precipitation and snow information, and sometimes antecedent runoff information depending on basin), and (b) default plus "teleconnections". For set (a), candidate information arrives on a station-specific basis. For set (b), individual climate signals antecedent to a given issue were treated as additional "information stations."

Information sets were compared on a basin-specific basis and at each forecast issue-stage during 1962-2000 for MP basins and 1961-2003 for PN basins. (Note: these periods are different than periods considered during signal selection in Part I.) For each simulation, the evaluation is conducted using a combinatorial implementation of a three-step process: (1) translate intercorrelated station variables into uncorrelated variables via principal component analysis (PCA) with noise truncation; (2) implement forecast model construction (approach described in next section) in a leave-one-out cross-validation framework to produce a time-series of validation forecasts; and (3) compute information metrics based on the validation forecasts. The combinatorial aspect involves implementing the three-step process using all subsets of the intercorrelated station variables antecedent to the issue being considered. This indicates the variable subset leading to optimal validation success, and leads to a filtering of signals selected from Part I. For shorter-lead forecasting, the analysis wrapper is first implemented with set (a). This leads to identification of best subset (a), which is then held constant in the set (b) wrapper analysis to identify the best subset of teleconnection signals to add to best subset (a).

Approach Details – Shorter-Lead Issues: Stepwise regression on the preprocessed PC's was used to construct the shorter-lead forecast models. This approach is similar to the model construction methodology used by several Western U.S. forecast centers (**Garen 1992**) (e.g., NWS CNRFC, NRCS National Water and Climate Center). Validation forecasts produced in the leave-one-out cross-validation framework were used to compute r^2 and other metrics.

Approach Details – Longer-Lead Issues: Discriminant Analysis (**Wilks 1995**) on the preprocessed PC's was used to construct the longer-lead forecast models. Discriminant Analysis differs from the stepwise regression analysis in that the outcome forecast is a set of probabilities assigned to several pre-defined runoff categories (e.g. historical quartiles) rather than a precise estimate of the expected outcome. The technique has been demonstrated for various hydroclimatic forecast settings (**Ward and Folland 1991, Piechota et al. 2001, Mason and Mimmack 2002**). For this application, each longer-lead model was developed to forecast probabilities of occurrence for historically-based quartiles. Validation skill for the categorical forecasts was assessed using several metrics, including Heidke Skill and Ranked Probability Skill (**Wilks 1995**). Each skill was scored relative to approximate climatology. Categorical forecasts were also converted into continuous distribution estimates using a resampling-with-replacement technique defined by the following five-step process: (a) use predictor information and forecast model to compute categorical probabilities for a given validation year, (b) generate a "forecast-weighted" runoff dataset by resampling historical category-sets at a ratio equal to the ratio of categorical forecast probabilities, (c) pool and sort the resampled-historical data, (d) associate sorted data with rank-probability plotting position, and (e) extract runoff values from the forecast distribution at exceedences of interest (e.g., 10%, 50%, 90% exceedence).

EXAMPLE RESULTS

Results are presented for two basins considered in the analysis: (1) Yakima River at Parker Road Crossing (Washington); and (2) Truckee River at Farad (California/Nevada). For the Yakima Basin, signals were selected from both North Pacific atmosphere conditions and ENSO-index conditions (not presented here). Seven North Pacific signals were ultimately selected for use in April-July runoff forecast models **(Table 1)**, with the collective locations of antecedent signals at each issue lead-time shown on **Figure 1**. Interpretation of signal search and evaluation results (**Figure 2**) suggests that above-average Yakima runoff is generally preceded by: (1) compressed Summer and Autumn $\Delta_{500\text{-}700}$ in the West Central Pacific, (2) expanded Summer $\Delta_{500\text{-}700}$ in the vicinity of Siberia/China, (3) expanded Summer Z_{700} northeast of Hawaii, (4) expanded Autumn Z_{700} over the Aleutians, compressed Autumn $\Delta_{500\text{-}700}$ offshore from British Columbia, (5) and *contracting* Summer-to-Autumn $\Delta_{500\text{-}700}$ over the West Central Pacific. Signal conditions associated with above-average Yakima runoff are also related to coincidental cooler-to-warmer equatorial Pacific SST from East to West (resembling ENSO "cool" phase conditions), and increased southerly atmospheric flow over the Yakima Basin during Dec-Feb (results not shown).

For the Truckee Basin three signals from North Pacific atmospheric conditions were selected for use in April-July runoff forecast models **(Table 1)**, with the collective locations of antecedent signals at each issue lead-time shown on **Figure 1**. Interpretation of signal search and evaluation results (**Figure 2**) suggests that above-average Truckee runoff is preceded by (1) compressed-but-*expanding* Summer $\Delta_{500mb\text{-}700mb}$ along China's East Coast, and (2) compressed Autumn Z_{700mb} over Northern California and Southern Oregon **(Figure 2)**. Signal conditions associated with above-average Yakima runoff are consistently related to coincident cooler Northwest Pacific SST (**Figure 2**), and increased southerly atmospheric flow over the Truckee Basin during Dec-Feb (results not shown).

Table 1 North Pacific Atmosphere Signal Selections and Correlation with Apr-Jul Runoff

Signal Name:	**Correlation**	**Informs Forecasts Issued in:**				
		Oct	Nov	Dec	Jan	Feb
Yakima Basin at Parker Road Crossing, Washington						
Z700Lat32.5-27.5Lon212.5-222.5Jun-JulMean	0.47	x	x	x	x	x
Z700Lat52.5-47.5Lon172.5-182.5Aug-OctMean	0.54		x	x		x
Z700Lat52.5-47.5Lon177.5-187.5Aug-NovMean	0.52				x	x
Z700Lat27.5-22.5Lon147.5-157.5Jun-NovTend	-0.46			x		
Diff500700Lat47.5-42.5Lon117.5-127.5Jul-AugMean	0.46	x				
Diff500700Lat27.5-22.5Lon142.5-152.5Jul-SepMean	-0.52	x	x			
Diff500700Lat27.5-22.5Lon152.5-162.5Sep-OctMean	0.52			x	x	
Diff500700Lat52.5-47.5Lon222.5-232.5Oct-NovMean	-0.45			x	x	
Truckee Basin at Farad, California						
Z700Lat42.5-37.5Lon232.5-242.5Oct-NovMean	-0.45				x	x
Diff500700Lat32.5-27.5Lon117.5-127.5Jun-JulMean	-0.48				x	x
Diff500700Lat37.5-32.5Lon117.5-127.5Jun-AugTend	0.52	x	X	x	x	x

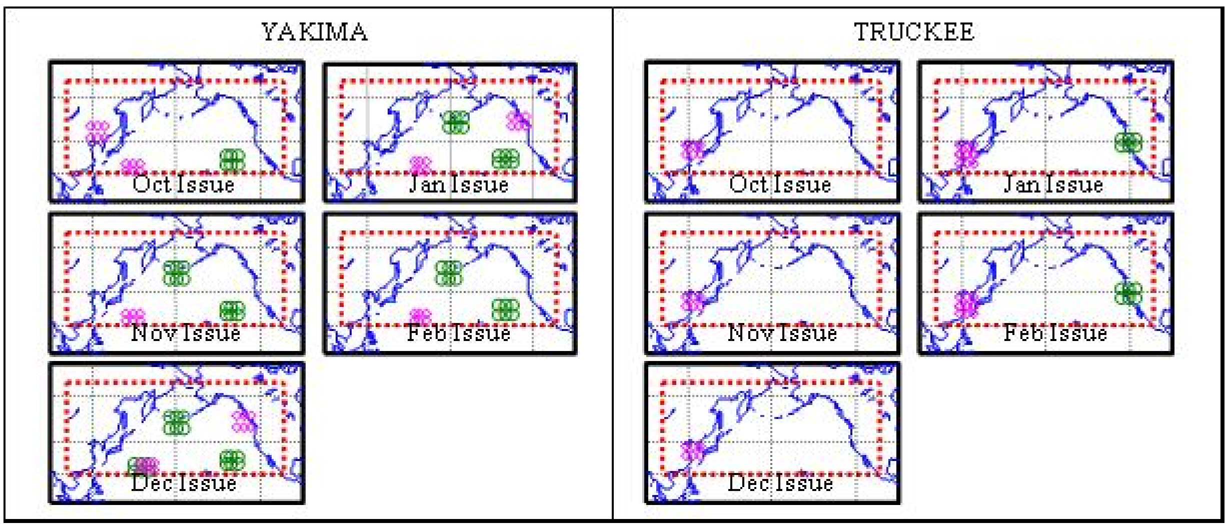


Figure 1 North Pacific Atmosphere Signal Selections, by Location and Antecedent to Forecast Issue-Month ($\Delta_{500\text{-}700}$: magenta diamond, Z_{700}: green circle)

Table 2 North Pacific Atmosphere Signal Effects on Current-Lead Forecast Performance

Issue Month	Validation r^2		RMSE of the Calibration Estimates (TAF)			Median April-July runoff, TAF
	No Tele.	With Tele.	No Tele.	With Tele.	Reduction	
Yakima Basin at Parker Road Crossing, Washington						
Jan	0.61	0.72	376	319	57	1710
Feb	0.82	0.85	256	230	26	
Truckee Basin at Farad, California						
Jan	0.28	0.45	137	119	18	267
Feb	0.55	0.64	107	95	12	

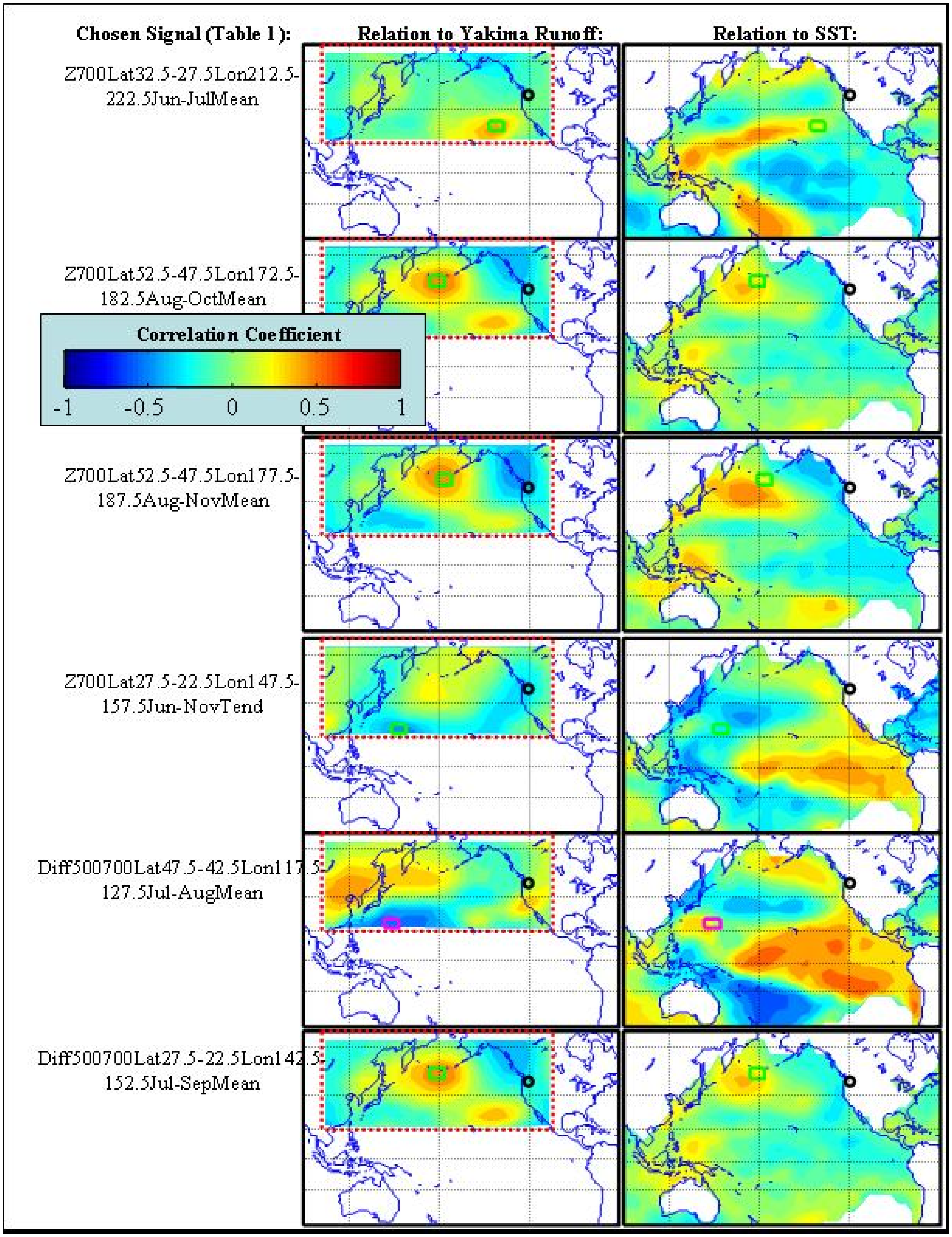
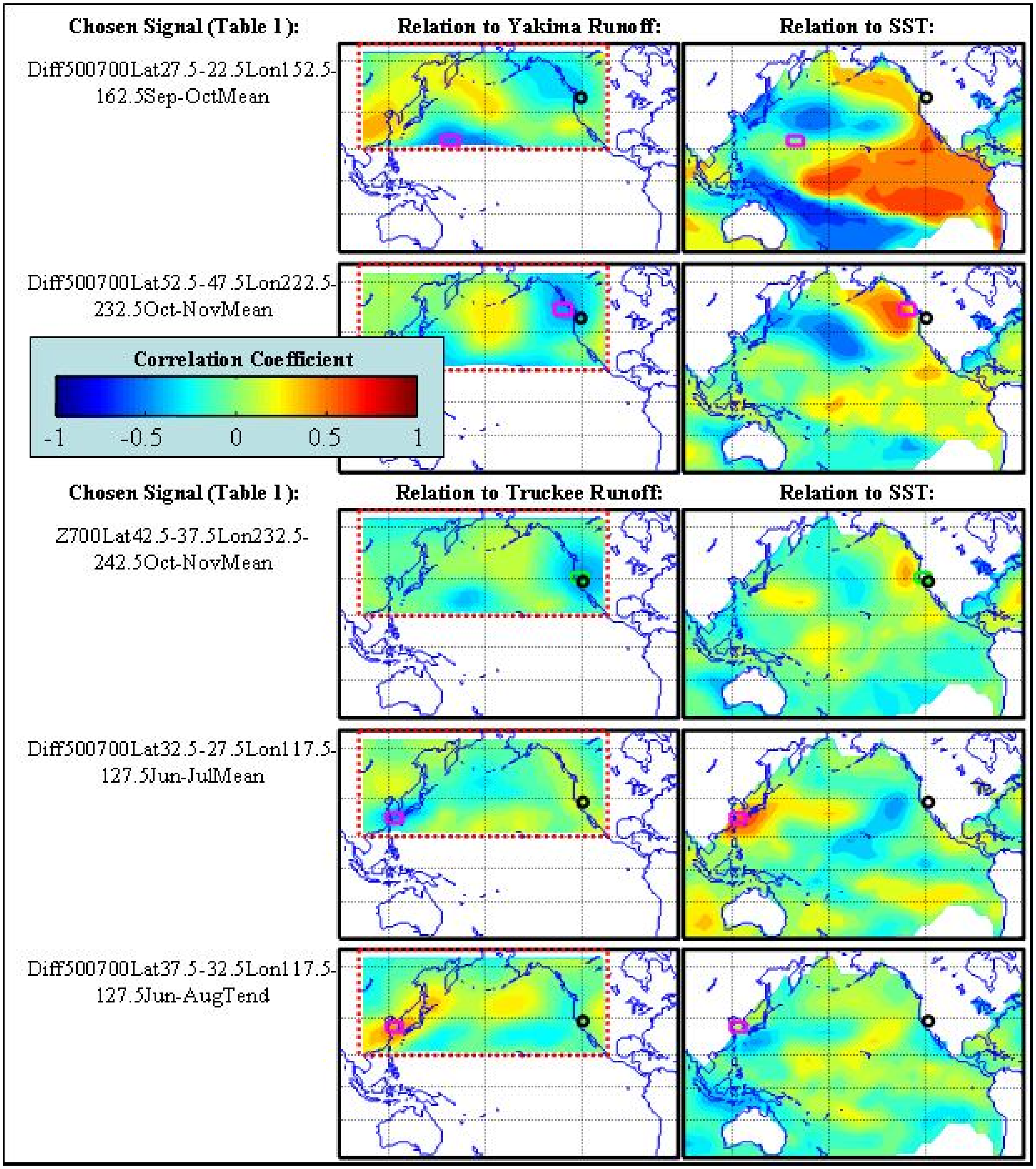


Figure 2 North Pacific Atmosphere Signal Relations to Target-Basin Runoff and Pacific-Basin SST (magenta/green boxes are signal locations, black circles are basin locations)



(Figure 2 Continued)

Table 2 summarizes the effects of using the **Table 1** signals to complement the January and February forecast information sets used in current operational forecasting by Reclamation PN-6200 and NWS CNRFC. Inclusion of teleconnection information sets was found to reduce calibration root-mean-square error by 10 to 15% for both basins and issues. Improvement was found for validation r-square for both the January and February issues, with most notable improvement found at the January issue. The evaluation also considered using antecedent signals to improve March and April forecasts (results not shown), but results showed that the signals offered little additional information value compared to the information already provided by season-to-date precipitation and snow conditions at those issues.

Table 3 summarizes validation results of using the **Table 1** signals to construct October, November and December forecast models of April-July runoff. Categorical forecasting skill was measured relative to climatology-based forecasting using two metrics: Heidke Skill and Ranked Probability Skill. There appears to be skill at each lead-time for each basin. Converted categorical-to-distribution forecasts were evaluated to determine whether the distribution's implied uncertainty consistently relates to the actual outcome. This evaluation was based on comparing the time series of validation forecast 80% uncertainty intervals (i.e. the difference between 10% and 90% exceedence estimates) and time series of actual outcomes. Results show the expected result where the forecast 80% uncertainty interval contained the actual outcome in nearly 80% of the validation cases for each basin-issue model. This implies that the forecast distribution exhibits an expected uncertainty-outcome relationship, and that the signals-based forecast reliability is consistent with the climatology-based forecast reliability. In addition to not harming forecast reliability, the signals information also affords significant reductions in the 80% uncertainty interval as early as October for both basins, and more aggressive anticipation of 90% exceedence estimate at the October, November, and December lead-times.

Table 3 North Pacific Atmosphere Signal Effects on Longer-Lead Forecast Performance

Issue Month	**Category Skill**		**Assessing Validation Forecast Distributions**		
	Heidke	**Rank Prob.**	**Frequency of Actual in Forecast 80% Uncertainty Interval**	**Mean Reduction in 80% Uncertainty Interval: No Info. minus Forecast (TAF)**	**Mean Chg. in 90% Exc: Forecast minus No Info. (TAF)**
Yakima Basin at Parker Road Crossing, Washington, 1961-2003 Analysis					
Oct	0.44	0.14	79%	266	173
Nov	0.41	0.22	79%	522	229
Dec	0.47	0.25	79%	580	309
Truckee Basin at Farad, California, 1962-2000 Analysis					
Oct-Dec	0.14	0.06	77%	103	14

Comparing results by basin, the Yakima signal selections appear to offer more convincing skill improvements than the Truckee signals selections. Perhaps the relative skill improvements are related to the physical interpretations of the teleconnection signals. The North Pacific signals appear to be secondary expressions of the ENSO as utilized in the Yakima (**Figure 2**). In contrast, the physical interpretation of Truckee signal selections is more difficult given that the selected atmospheric signals appear to be related to North Pacific sea surface temperatures that do not seem related to ENSO or other known phenomena. More physical evaluation is required to gain understanding on the Truckee signals.

Finally, it was mentioned that there were also ENSO-Index selections for the Yakima Basin. These selections were based on SOI, MEI, Nino4 indices. Their arrival, like the North Pacific atmosphere signals, also occurred from Summer to Winter prior to the April-July period. Joint consideration of mean ENSO-index conditions ($MEI_{Aug\text{-}Sep}$ and $SOI_{Jun\text{-}Sep}$) with signals listed in Table 1 was found to slightly improve skill for October-issue (results not shown); skill improvements were not found for November- and December-issues.

RESEARCH TO OPERATIONS

The analysis showed that integration of the selected North Pacific atmospheric signals with season-to-date local precipitation and snowpack information can improve January and February forecasts of April-July runoff in nearly all of the 17 focus basins. Longer-lead forecasting at October, November, and December lead-times was also found to have skill over climatology in most of the focus basins.

The beneficiaries of improved April-July runoff forecasts at January and February lead-times would include current clients of NRCS, NWS, Reclamation PN-6200, and CA DWR forecasts (e.g., Reclamation's Mid-Pacific and Pacific Northwest water project operators, CA State Water Project operators). The beneficiaries of longer-lead forecasts are less obvious due to the current absence of both Autumn-lead forecasts products and decision-processes based on them. That said, it seems reasonable to expect relatively risk-neutral decision-makers (e.g., long-term water and energy wholesalers) to benefit from this longer-lead information that reduces uncertainty relative to climatology.

To set up potential transition of results from research to operations, this project was designed as collaboration between project investigators and several operational forecasting centers (NRCS, NWS CNRFC, Reclamation PN-6200, and CA DWR). The goals were to keep these centers informed on the research approach and progress, gain their feedback methods and results significance, and involve them on determining follow-up steps. On the latter, feedback from each center was used to frame the development of an experimental forecasting tool that would allow further evaluation of data streams and forecast models at the centers. The tool performs basic functions: (a) import daily/monthly signal data from NCEP/NCAR Reanalysis data at CDC, (b) manage and export seasonal signal data to the forecast center's data systems, (c) archive and generate Autumn lead-time forecasts of April-July runoff, and (d) present metadata on North Pacific signals for October- through February-issues similar to graphics shown on **Figure 2**. The tool was deployed to each center in August 2005. Evaluation is on-going at the centers. Initial focus will be placed on using the tool's climate signal information to improve January and February forecast equations. Secondary focus will be placed on gaining familiarity with the longer-lead models and deciding on how to potentially map these experimental products with prospective clients. Pending results and incorporated tool-improvements based on this evaluation, a public-release version is expected to be made available sometime in 2006.

REFERENCES

Clark, M.P., and Serreze, M.C. (1999). "Snowfall responses over the USA to phase and amplitude variations in the tropospheric wavetrain," Proc. IUGG 99 Symposium HS2, Birmingham, AL.

Garen, D.C. (1992). "Improved techniques in regression-based streamflow volume forecasting" ASCE Journal of Water Resources Planning and Management, 118(6), 654-670.

Grantz K., Rajagopalan, B., Clark, M., and Zagona E. (2005). "A Technique for Incorporating Large-Scale Climate Information in Basin-Scale Ensemble Streamflow Forecasts," Water Resources Research (In Press).

Hamlet, A. F., Huppert, D., and Lettenmaier D. P. (2002). "Economic value of long-lead streamflow forecasts for Columbia River hydropower," ASCE Journal of Water Resources Planning and Management, 128(2), 91-101.

Redmond, K.T., and Koch, R.W. (1991). "Surface climate and streamflow variability in the western United States and their relationship to large scale circulation indices," Water Resources Research, 27, 2381-2399.

Souza, F.A., and Lall, U. (2003). "Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: applications of a multivariate, semi-parametric algorithm," Water Resources Research, 39, 1307-1320.

Wilks, D.S. (1995). Statistical Methods in the Atmospheric Sciences. Academic Press, San Diego, CA, 467pp.

Yarnal, B., and Diaz, H. F. (1986). "Relationships between extremes of the Southern Oscillation and the winter climate of the Anglo-American Pacific coast," Journal of Climatology, 6, 197-219.